先端人工知能論Ⅱ

高度な画像認識 (Part 2)
～物体検出～

東京大学 大学院情報理工学系研究科
創造情報学専攻　講師
中山 英樹

一般物体認識の主要なタスク

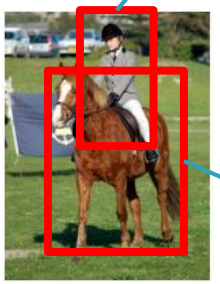
易

難

- Categorization (カテゴリ識別)
  - 映ってる物体の名称を答える
  - 物体の位置を答える必要はない

horse
human

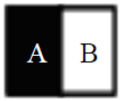
- Detection (物体検出)
  - 矩形で物体の位置を切り出す



- Semantic Segmentation
  - ピクセルレベルで物体領域を認識

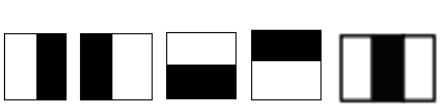
深層学習以前(1)

- Haar-like 特徴による顔検出 [Viola & Jones, CVPR '01]



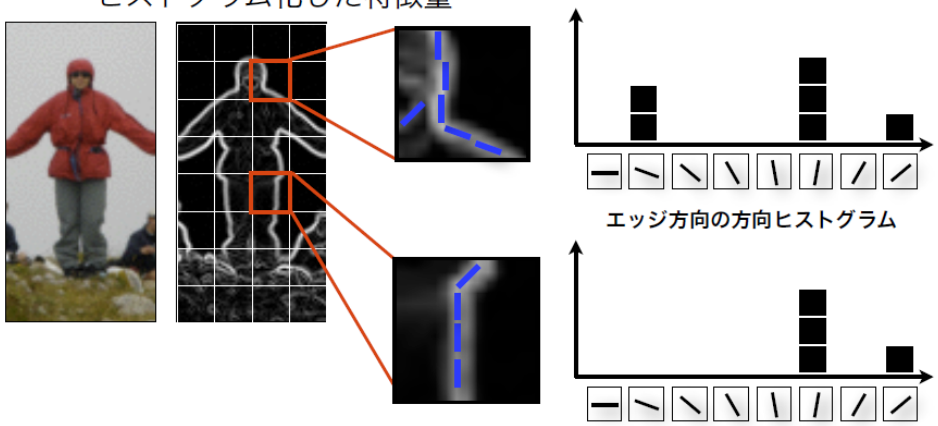
$$H(r1,r2) = S(r1) - S(r2)$$

領域Aの平均輝度　　領域Bの平均輝度


Haar-likeパターン


顔の特徴を捉えることが可能

- HOG特徴 [Dalal, CVPR '05]
  - 局所領域におけるエッジ(勾配)の方向を
    ヒストグラム化した特徴量

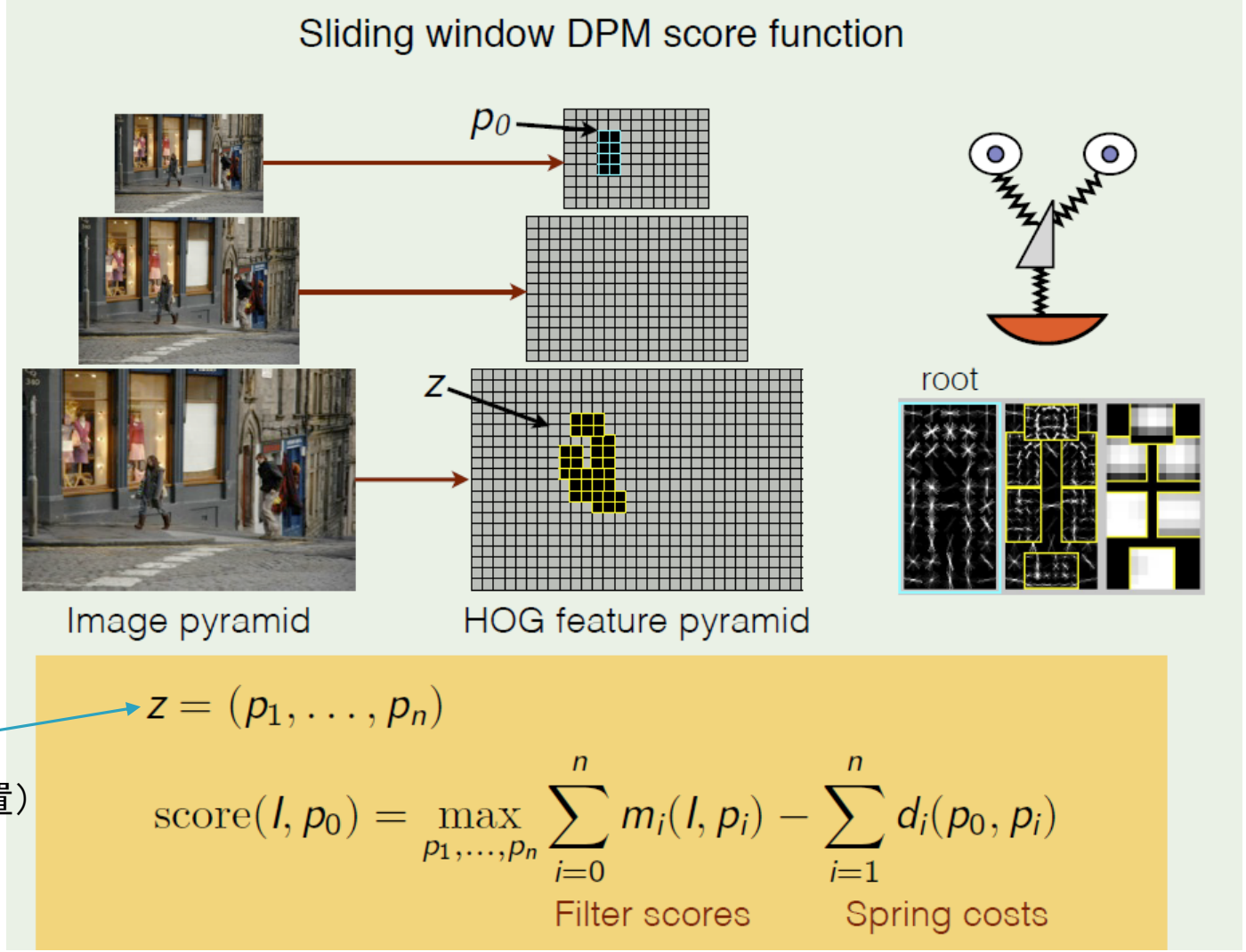

エッジ方向の方向ヒストグラム

※中部大　藤吉先生の
スライドより引用

深層学習以前(2)

Slide courtesy of Ross Girshick
http://vision.stanford.edu/teaching/cs231b_spring1213/slides/dpm-slides-ross-girshick.pdf

- Deformable part model [Felzenszwalb+, CVPR'08]



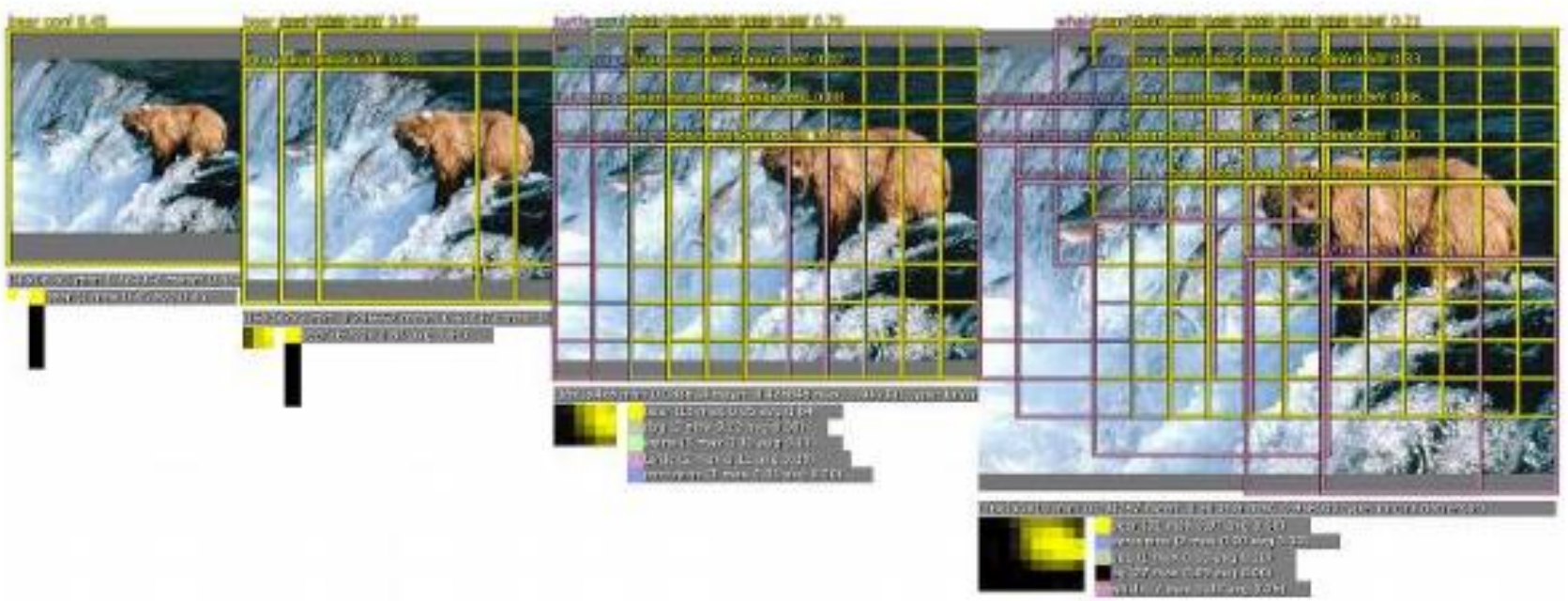
潜在変数
（各パーツの位置）

物体検出の難しさ

- 候補となる領域の多い
  - ナイーブなsliding windowは計算量が爆発
- CNNの場合、プーリングにより解像度が落ちる
  - 高精度な位置推定が本質的に困難

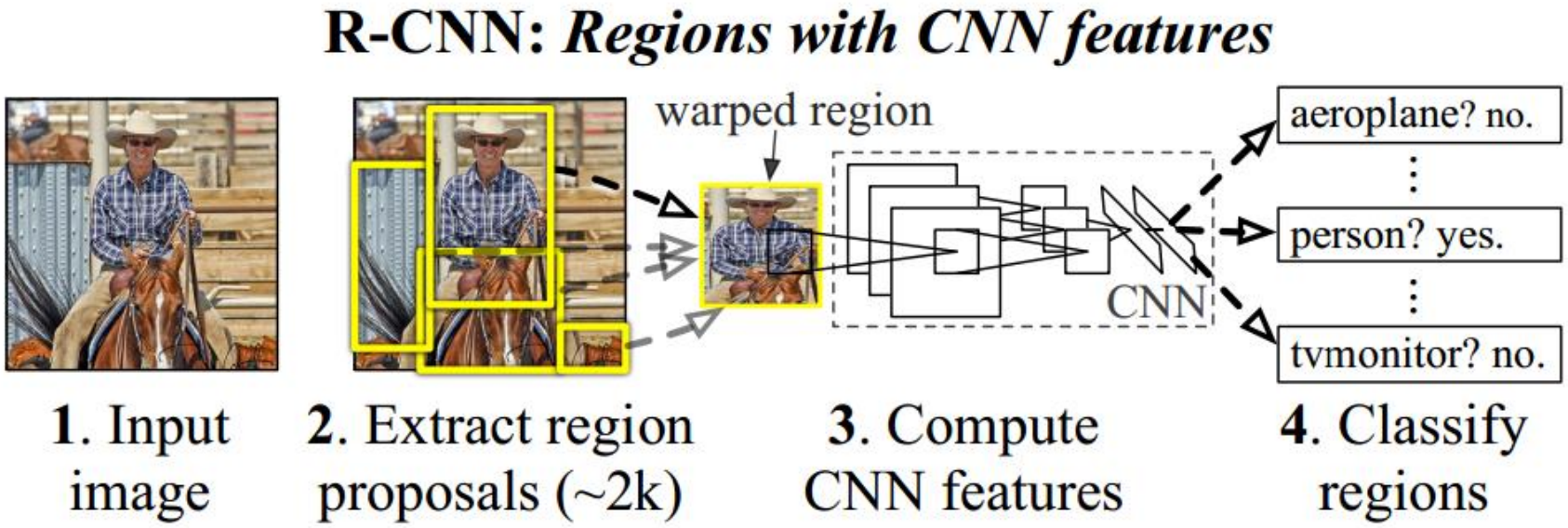
Detection by Regression

- OverFeat [Sermanet et al., ILSVRC'13]
  - 最終畳み込み層（アップサンプリング込み）の各グリッドでクラス識別+物体領域の座標を回帰
  - 計算はほとんどの部分が共有できるので高効率



Detection by Object Region Proposals

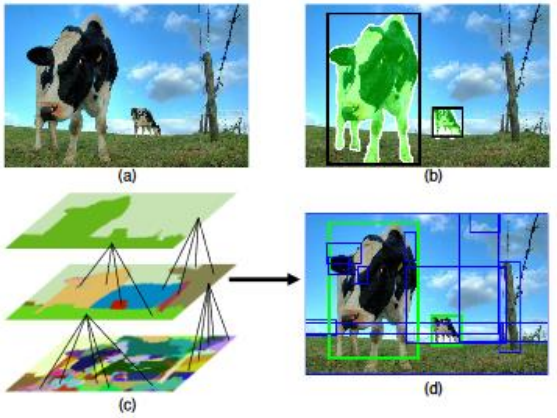
- R-CNN [Girshick et al., CVPR'2014]
  - 物体の領域候補を多数抽出（これ自体は別手法）
  - 無理やり領域を正規化し、CNNで特徴抽出
  - SVMで各領域を識別



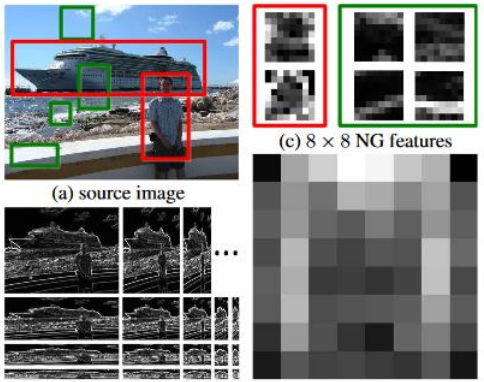
Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation", In Proc. CVPR, 2014.

Region proposal (Region of interest)

- Selective search
  - [van de Sande et al., CVPR'11]
  - セグメンテーションベース



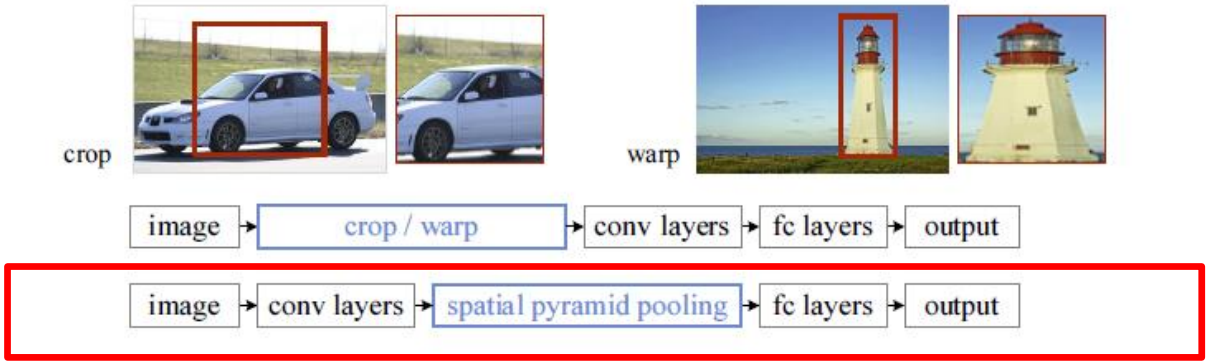
- Bing
  - [Cheng et al., CVPR'14]
  - 入力画像を変換して候補領域を探索
  - とても速い

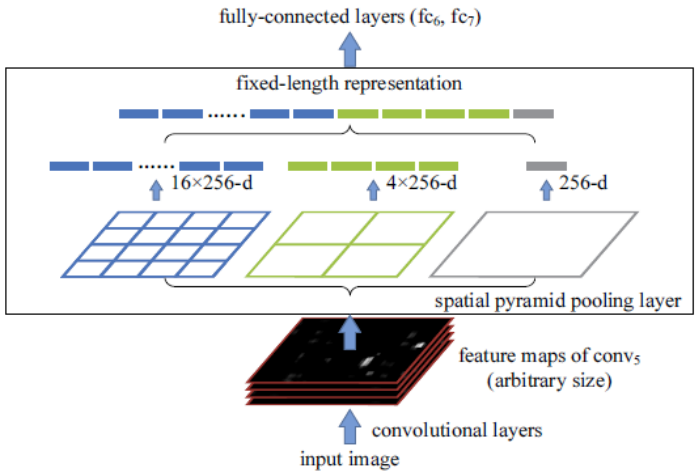


- Multibox
  - [Erhan et al., CVPR'14]
  - Deep learningで候補領域を生成

Region warping はどうにかならないか？

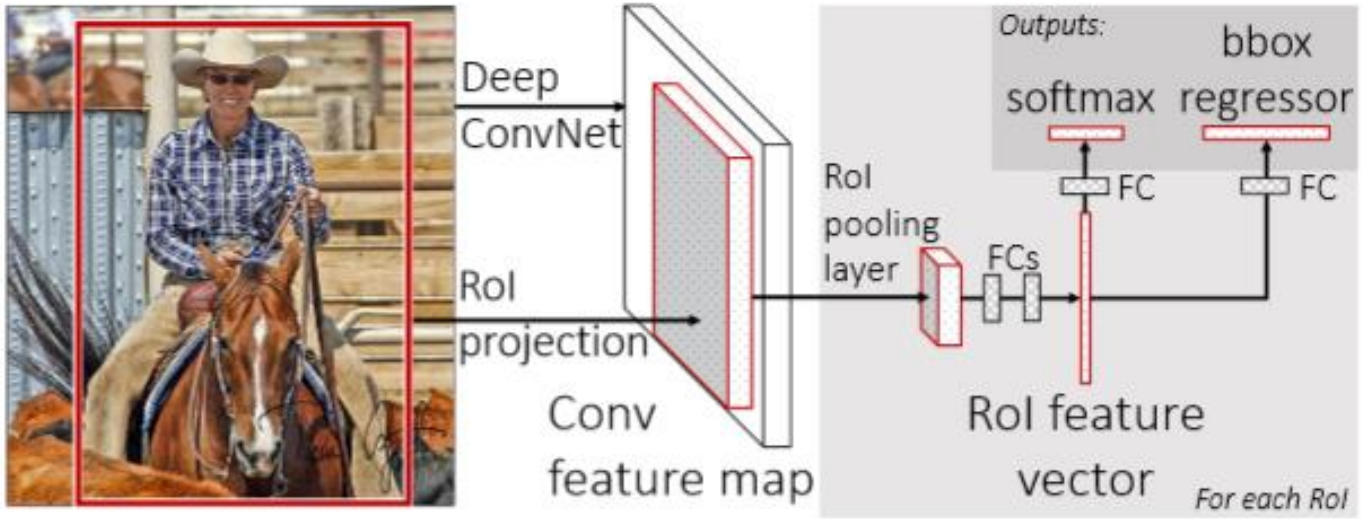
- Spatial pyramid pooling [He+, ECCV'14]
  - 入力に対する、畳み込み層の相対位置でのプーリングでOK





R-CNNの拡張(1)

- Fast R-CNN [Girshick, ICCV'15]
  - Region warping の代わりにSPPを導入
  - 識別速度を数百倍高速化
  - 学習の効率化



R-CNNの拡張(2)

- Faster R-CNN [Ren et al., NIPS'15]
  - 物体候補領域自体をNNで生成
  - 識別層までend-to-endで学習
  - より少ない候補領域から
    高精度な認識

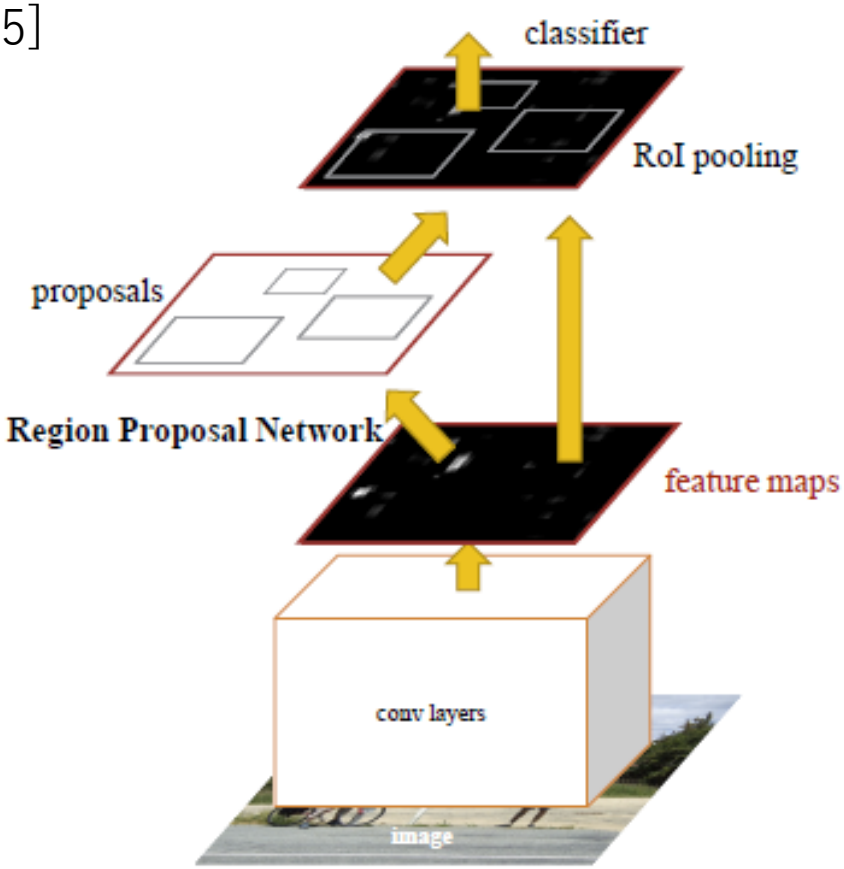
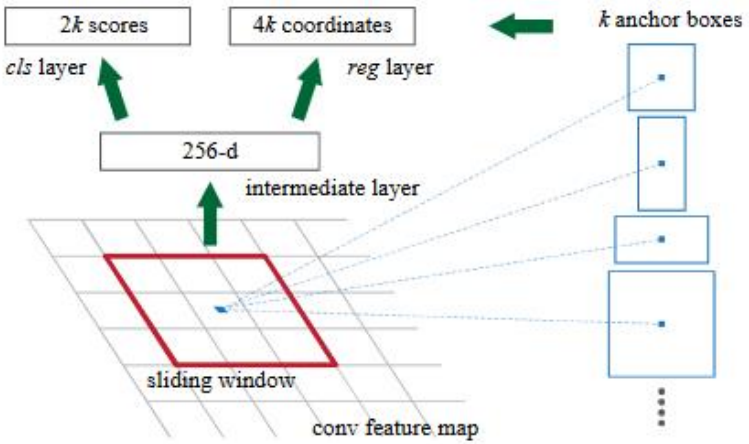


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

Region Proposal Network (RPN)

- 特徴マップの各受容野(sliding window)で物体領域を予測
  - 各領域の位置・大きさを回帰、物体/非物体の識別
    - **k候補領域→(4+2) x kの出力数**
  - RPN部のパラメータは全領域で共有（位置不変）
    - **実際には1x1の畳み込みレイヤとして実装可能**
    - **Multiboxと比較して大幅なパラメータ数の削減、高速化**



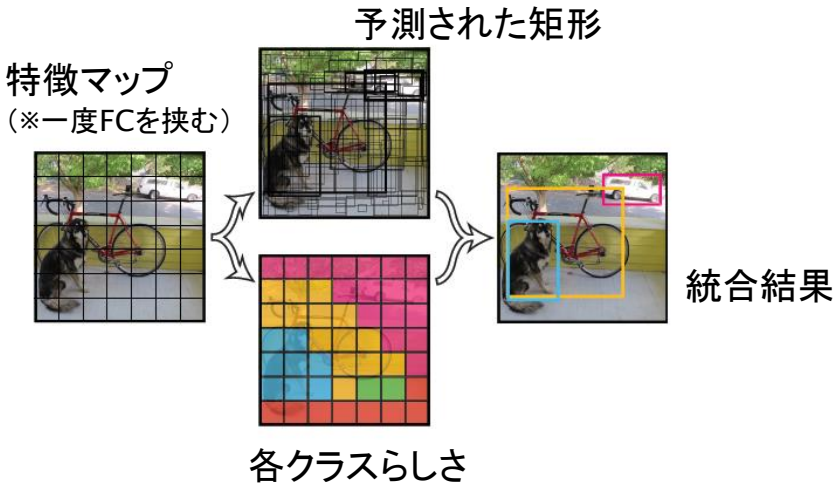
余談：発表時系列

- R-CNN arXiv初版　2013年11月
- R-CNN 発表@CVPR'14　2014年6月
- Fast R-CNN arXiv初版　2015年4月
- Faster R-CNN arXiv初版　2015年6月
- Faster R-CNN 発表@NIPS'15　2015年12月
- Fast R-CNN 発表@ICCV'15　2015年12月

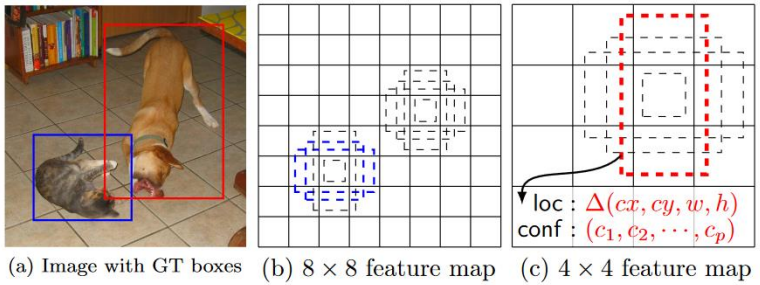
Region proposal もいらないのでは…？

- Single-shot アーキテクチャの台頭
  - 直接的に各物体クラスらしさを推定 (bounding box回帰と同時)

- YOLO (you only look once) [Redmon et al., CVPR'16]



- SSD (single-shot multibox detector) [Liu et al., ECCV'16]

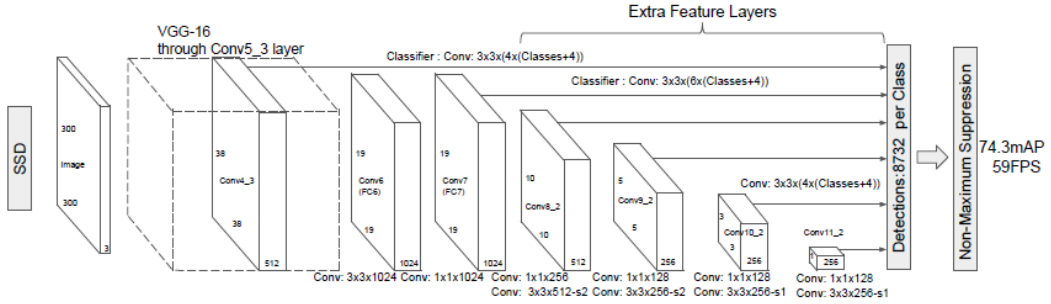


(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

※複数の特徴マップを利用するための工夫がされている

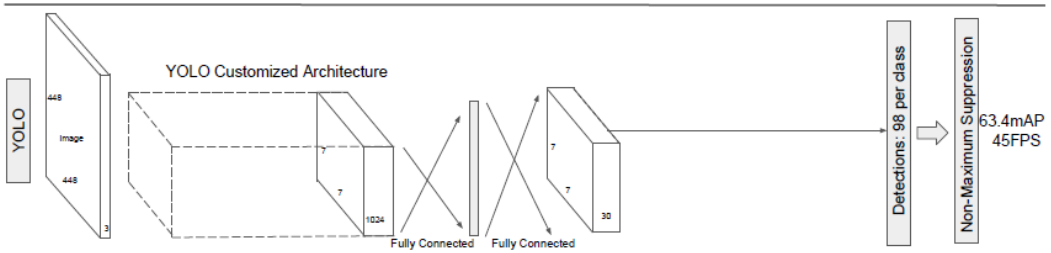
Single-shot系アーキテクチャの比較

SSD



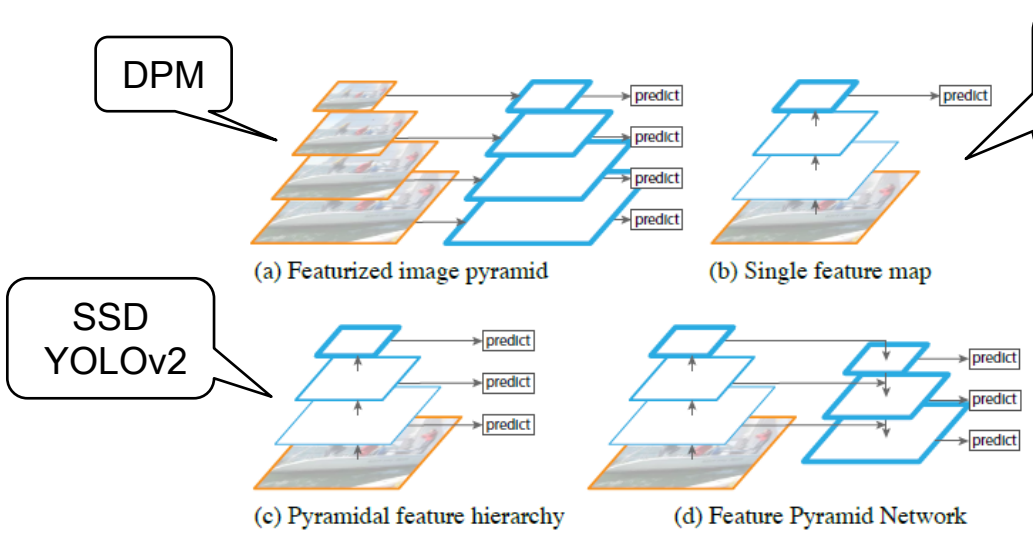
74.3mAP
59FPS

YOLOv1



63.4mAP
45FPS

[Liu et al., ECCV'16]



[Lin et al., CVPR'17]

Redmon and Farhadi, “YOLO9000: Better, Faster, Stronger”, In Proc. CVPR, 2017.

性能比較

	YOLO							YOLOv2
batch norm?		✓	✓	✓	✓	✓	✓	✓
hi-res classifier?			✓	✓	✓	✓	✓	✓
convolutional?				✓	✓	✓	✓	✓
anchor boxes?				✓	✓			
new network?					✓	✓	✓	✓
dimension priors?						✓	✓	✓
location prediction?						✓	✓	✓
passthrough?							✓	✓
multi-scale?								✓
hi-res detector?								✓
VOC2007 mAP	63.4	65.8	69.5	69.2	69.6	74.4	75.4	76.8	**78.6**

Detection Frameworks	Train	mAP	FPS
Fast R-CNN [5]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[15]	2007+2012	73.2	7
Faster R-CNN ResNet[6]	2007+2012	76.4	5
YOLO [14]	2007+2012	63.4	45
SSD300 [11]	2007+2012	74.3	46
SSD500 [11]	2007+2012	76.8	19
YOLOv2 288 × 288	2007+2012	69.0	91
YOLOv2 352 × 352	2007+2012	73.7	81
YOLOv2 416 × 416	2007+2012	76.8	67
YOLOv2 480 × 480	2007+2012	77.8	59
YOLOv2 544 × 544	2007+2012	**78.6**	40

Region proposalとは何だったのか？（雑感）

- Single-shot でうまくいくようになった
  - 結局OverFeatと同じアプローチへ戻る
- Region proposalは（広い意味で）カリキュラム学習の一つ
  - 難しい問題をサブタスクへ分割するとうまくいくことが多い
  - データ量や、周辺技術の進展で無意味になることも
- 深層学習界隈ではよくある展開
  - （逆に復活することも…？）

演習

- Faster-RCNN, SSD (ChainerCV)
- YOLOv2のChainerコードの解説・実行